

Towards Cancer Hybrid Automata

Loes Olde Loohuis
CUNY The Graduate Center
Department of Computer Science
l.olde@loohuis@gmail.com

Andreas Witzel
NYU
Courant Institute
awitzel@nyu.edu

Bud Mishra
NYU
Courant Institute
mishra@nyu.edu

This paper introduces Cancer Hybrid Automata (CHAs), a formalism to model the progression of cancers through discrete phenotypes. The classification of cancer progression using discrete states like stages and hallmarks has become common in the biology literature, but primarily as an organizing principle, and not as an executable formalism. The precise computational model developed here aims to exploit this untapped potential, namely, through automatic verification of progression models (e.g., consistency, causal connections, etc.), classification of unreachable or unstable states and computer-generated (individualized or universal) therapy plans. The paper builds on a *phenomenological* approach, and as such does not need to assume a model for the biochemistry of the underlying natural progression. Rather, it abstractly models transition timings between states as well as the effects of drugs and clinical tests, and thus allows formalization of temporal statements about the progression as well as notions of timed therapies. The model proposed here is ultimately based on *hybrid automata*, and we show how existing controller synthesis algorithms can be generalized to CHA models, so that therapies can be generated automatically. Throughout this paper we use cancer hallmarks to represent the discrete states through which cancer progresses, but other notions of discretely or continuously varying state formalisms could also be used to derive similar therapies.

1 Introduction

Cancer is generally thought of as a *progressive disease* – in particular, a disease which exhibits certain discernible cancer phenotypes (modeled as a finite set of *discrete* states), through which it progresses towards a terminal phenotype (e.g., metastasis).

Among other theories, this view is reflected in the so-called *hallmarks of cancer* proposed by Hanahan and Weinberg [8], and it has become one of the predominant ways of thinking about cancer, solidified through many further publications and experiments. A recent article by the same authors [9] reviews and consolidates the new insights of the last decade. Similar models have also been explored by a mechanistic agent-based simulation in [1].

According to the model proposed by Hanahan and Weinberg, tumors must necessarily acquire certain “intermediate” hallmarks culminating in the “final” hallmarks of tissue invasion and metastasis. As the authors write,

Simply depicted, certain mutant genotypes confer selective advantage on subclones of cells, enabling their outgrowth and eventual dominance in a local tissue environment. Accordingly, multistep tumor progression can be portrayed as a succession of clonal expansions, each of which is triggered by the chance acquisition of an enabling mutant genotype. [9, p. 658]

The current list of cancer hallmarks includes the abilities to reproduce autonomously, to ignore anti-growth signals, or to signal for formation of new blood vessels, as well as handful of other phenotypes. Hallmarks can be obtained in various different orders, but not every order is viable. Intuitively, a hallmark can be acquired by a dominant sub-population of cells if it conveys a selective advantage compared to the

other phenotypes acquired in that population. For example, in a wildly growing cluster of cells, the ability to signal for new blood supply, and thus nutrients, oxygen, and waste disposal, will allow the respective sub-population to outgrow the others.

Most hallmarks are acquired through mutations (point mutations, copy number changes or epigenetic modifications) of very specific sets of oncogenes and tumor suppressor genes. Thus, many of the targeted drugs, administered individually or combinatorially in a cocktail, which have been developed in recent years, aim to influence the function of the products of these genes [16] and thus cancer's evolution from specific hallmarks. For example, the vascular endothelial growth factor (VEGF) signals for creation of new blood vessels (*angiogenesis*), and the drug Avastin inhibits the associated signaling pathway, thus preventing growing tumors from obtaining the needed blood supply. While current therapies target only the observed hallmark at any instant, they rarely take into account the potential hallmarks that may evolve in the future and the temporal structure of the underlying evolution. By connecting therapy design to the theory of supervisory control of hybrid automata, we aim to build a framework for better therapy design (e.g., that avoids drug-resistance, exploits synthetic lethality, oncogene addiction, etc., and avoids undesirable side-effects on other organs).

In this view of cancer, its progression through hallmarks and therapy bears a striking resemblance to formal models of state-transition machines in computer science.

In this paper, we first present a logical framework called *Cancer Hybrid Automaton* (CHA) that allows us to formally capture cancer progression through accumulation of successive discrete states. States in CHA models represent states of the progression, and directed edges among pairs of states define possible progression paths. Drugs can then be thought of as inhibiting or prolonging specific transitions in the automaton. We then show how this approach enables us to formally describe cancer progression, automatically verify/model-check its temporal properties, and manipulate its evolution to satisfy certain therapeutic goals.

We illustrate our approach through a highly simplified running example of a cancer hybrid automaton in which states represent hallmarks, and progression paths represent successive hallmark acquisitions. However, the states of the automaton can represent any set of discrete states at varying levels of abstraction. Examples include stages of cancer, a set of affected pathways, and a set of specific genomic aberrations. By ignoring complex structures such as heterogeneity, geometry, circulating tumor cells, tumor growth dynamics, genomic instability at this point, we avoid obscuring the key ideas inherent to the therapy design algorithms. However, the framework is flexible enough to include such structures as well as detailed mechanistic models of the discrete states.

2 Overview

The rest of this paper is organized as follows. In section 3, we introduce a basic CHA formalism. In this section, a CHA is modeled as a *finite non-deterministic automaton*. The edges, representing transitions from one progression state (e.g hallmarks) to the next, are labeled with drugs that can inhibit the transition. A *therapy* is defined as a function that assigns a set of drugs to each finite progression history, or *run*. An execution of a therapy is defined as a run of the CHA that respects the therapy, that is, no transition of the execution is inhibited by the therapy. Our model includes costs by associating a cost vector with each state and each cocktail. Therapies may be selected by comparing costs of possible executions using a notion of Pareto dominance, in addition to the required qualitative properties specified in CTL.

In section 4 we extend the CHA framework to include real time. In this model, transitions take certain durations of time, and drugs can prolong (or stop) the transition process. This is modelled using a hybrid

automaton with multiple clocks¹. Clock constraints on the edges and clock invariants at the states restrict the possible progressions of the system. Multiple clocks are needed to allow for the scenario that a drug affects the transition to possible next states in different ways. Possible runs and therapies of a timed CHA now include the clock values. An extension of CTL, Timed CTL, is used to specify extended goals about the system.

In section 5, we discuss the problem of automatically generating therapies, i.e., controller synthesis for CHAs. For simple untimed CHAs this is a well-studied problem and algorithms exist. For timed CHAs, we show that if we allow only for control at discrete moments in time the problem is decidable for CTL goals.

Finally, section 6 concludes with a discussion of several possible extensions of our model, which will be addressed in the future work.

3 Cancer Hybrid Automata

A simple, intuitive example CHA is shown in fig. 1. It comprises the following hallmarks (see [8] for more details):

SSG: Self-sufficiency in growth signals. Roughly speaking, cells no longer depend on external growth-promoting signals, but grow autonomously. Usually, such a state is associated with a gain of function of an oncogene or a loss of function of a tumor suppressor gene.

IAG: Insensitivity to anti-growth signals. Cells with this hallmark continue to grow even in the presence of inhibiting signals. Usually, certain cell-cycle checkpoints are no longer properly regulated.

Ang: Sustained angiogenesis. This state enables a cancer cell to signal for the construction of blood vessels.

LRP: Limitless replicative potential. While most normal cells can only divide a certain number of times, cells with this hallmark can divide without limits. In this state, a cancer cell may upregulate telomerase to restore telomere lengths.

EvAp: Evading apoptosis. Normally, cells have a program for controlled cell-death, which is used to remove damaged or otherwise unwanted cells. This program is disabled in this hallmark, which allows cells with highly corrupted DNA to survive – thus facilitating cancer progression further.

M: Metastasis. This state enables cancer cells to spread from their original location to other parts of the body.

Various possible progressions through these hallmarks can be seen as transitions in the picture (note that this is a simplified and incomplete model). For example, Ang can be acquired after SSG and IAG. Moreover, as mentioned in section 1, if a growing tumor fails to acquire Ang, it may starve; in this case, a solid tumor is unable to grow further and attain the later hallmarks. For simplicity, it may be modeled as a transition to the normal state.

In this example, the therapy “give the drug Avastin whenever a state leading up to Ang is reached” will prevent the cancer from reaching M.

¹Hence the term *hybrid* in ‘cancer hybrid automaton’.

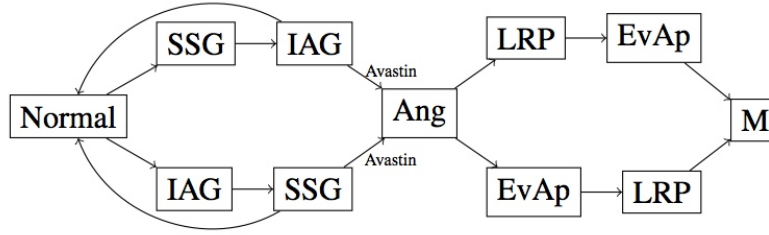


Figure 1: A simple CHA whose progression can be stalled by a VEGF-inhibitor such as Avastin.

3.1 Formal model

In the following, we start with a preliminary and simple formalization of the notions described above. We will successively extend the formal model in the later sections.

We assume a global set D of **drugs**.

Definition 0.1. A *Cancer Hybrid Automaton (CHA)* is a tuple

$$H = (V, E, v_0) ,$$

where

- V is a set of states,²
- $E \subseteq V \times 2^{\mathcal{D}} \times V$ is a set of directed edges labeled with sets of drugs, and
- $v_0 \in V$ is the initial state.

We usually omit v_0 and write just (V, E) .

Intuitively, an edge (v, D, v') represents a transition from state v to state v' that can be inhibited by any drug from the set $D \subseteq \mathcal{D}$. We allow several drugs to be given simultaneously and refer to such sets $C \subseteq \mathcal{D}$ of drugs as **cocktails**. Given a cocktail C , the edge $(v, D, v') \in E$ is *inhibited* by C if $C \cap D \neq \emptyset$. Given a state v and a cocktail C , v can transition to v' under C , in symbols $v \xrightarrow{C} v'$, if there is an edge (v, D, v') that is not inhibited by C . Note that we allow multiple edges (with different labels) between the same two states. To prevent a transition between two states, all edges connecting them need to be inhibited, which is why we need to consider cocktails rather than just single drugs. We assume that for every state v and every cocktail C there exists some state v' such that $v \xrightarrow{C} v'$ (possibly $v' = v$, these edges were omitted in fig. 1).

A **run** of a CHA $H = (V, E, v_0)$ is a sequence of transitions in E . Let $Runs(v, H)$ denote the set of runs that start in v . We write $Runs(H)$ for $Runs(v_0, H)$, and by $Runs_f(v, H)$ we denote the set of finite runs from $Runs(v, H)$.

We now formalize how it is possible to *interfere* with the progression of the system.

Definition 0.2. A **therapy** is a function $\pi : Runs_f(H) \rightarrow 2^{\mathcal{D}}$. A **possible execution** of π in H is a run

$$S = v_0 v_1 v_2 \dots ,$$

such that for each $i \geq 0$, $v_i \xrightarrow{\pi(S_i)} v_{i+1}$, where S_i denotes the initial segment of S up to step i .

²Strictly speaking, in the case of hallmarks, a state corresponds to a subset of hallmarks that have been acquired.

Definition 0.3. *Costs are given by the following (overloaded) function, for some finite dimension n :*

- $c : V \rightarrow \mathbb{R}_{\geq 0}^n$ specifying costs of states,
- $c : 2^{\mathcal{D}} \rightarrow \mathbb{R}_{\geq 0}^n$ specifying costs of cocktails.

Thus, both states and cocktails have costs assigned to them, represented as n -dimensional vectors. Dimensions may include toxicity of the drugs, monetary cost of the drugs, discomfort for the patient, etc.

The cost of a possible execution $S = v_0 v_1 v_2 \dots$ of a therapy π with discount factor $0 < \delta \leq 1$ is

$$c(S, \pi, H) = \sum_{i \geq 0} \delta^i (c(v_i) + c(\pi(S_i))) .$$

The set of possible costs of π for a CHA H is

$$c(\pi, H) = \{c(S, \pi, H) \mid S \text{ is possible execution of } \pi \text{ in } H\}.$$

Now that we have a definition of the set of possible costs of a therapy, we can compare different therapies with respect to their costs.

Definition 0.4. *A cost vector $x \in \mathbb{R}^n$ **Pareto-dominates** another vector $x' \in \mathbb{R}^n$, in symbols $x \prec x'$, iff for each $1 \leq \ell \leq n$ we have $x_\ell \leq x'_\ell$ and for some $1 \leq \ell \leq n$ we have $x_\ell < x'_\ell$.*

*A therapy π **Pareto-dominates** a therapy π' in a CHA H if for each $x \in c(\pi, H)$ and $x' \in c(\pi', H)$ we have $x \prec x'$. The set of **candidate therapies** for H is*

$$\Theta(H) = \{\pi \mid \pi \text{ is not Pareto-dominated in } H\} .$$

For the special case of 1-dimensional costs (or if there is a function to aggregate cost vectors into single numbers), the set of candidate therapies is the set of therapies whose best-case cost is not higher than some other therapy's worst-case cost.

This definition of a set of candidate therapies is a very conservative one, in that it includes any therapy that is not overtly worse than some other therapy. There are different possibilities for defining the set of candidate therapies, or for pruning the set further. Examples of such strategies for pruning the set further include *maximin*, i.e., choosing those strategies that lead to the best worst-case outcome, or *maximax*, i.e., choosing those strategies that lead to the best best-case outcome. However, making these decisions depends on the risk attitude of patient and doctor which may not be fully formalizable. Therefore we include all the potentially relevant therapies in the set of candidate therapies.

In order to be clinically applicable, a CHA model may need to be *personalized* for any given patient or cancer type. This personalization will result in families of CHAs, with different sets of candidate therapies. While we will not give full details here, we wish to describe one possible application for such richer models.

For families of automata, we can ask whether there are any *universal* therapies for all of the included automata. Such therapies can result in faster and cheaper treatments.

To be able to apply therapies across different automata, their domain must be the same. This requirement can be satisfied, for example, by considering CHAs that contain the same set of hallmarks, and therapies that either depend only on the current state, or that have the set of all sequences of states as domain. The following definition applies to therapies on such unified domains.

Definition 0.5. *Given a family \mathcal{H} of CHAs, the set of (**universal**) **candidate therapies** for \mathcal{H} is*

$$\Theta(\mathcal{H}) = \bigcap_{H \in \mathcal{H}} \Theta(H) .$$

A set θ of therapies **covers** \mathcal{H} if

$$\theta \cap \Theta(H) \neq \emptyset \text{ for all } H \in \mathcal{H} .$$

Note that if $\Theta(\mathcal{H}) \neq \emptyset$ then for each $\pi \in \Theta(\mathcal{H})$, $\{\pi\}$ covers \mathcal{H} .

3.2 Temporally extended goals: CTL

We have seen in the previous section that therapies can be compared according to their costs. Thus, the problem of finding the right therapy can be viewed as an optimization problem. It can, however, be necessary to have more detailed control over the therapeutic objectives. Simple reachability properties can be used as goals, such as “metastasis must never be reached”. For more expressivity we can use Computation Tree Logic (CTL) [4] to specify goals.

Example 1. The goal $AG\neg M$ states that metastasis is never reached. Another possible goal could be

$$AG(Ang \rightarrow AG\neg EvAp) .$$

This sentence means that whenever sustained angiogenesis is acquired, then at no point in the future the capability of evading apoptosis will be obtained.

One may be interested in checking properties of the CHA itself, without application of a therapy. This goal can be achieved by using CTL model checking (see, e.g., [5]). CTL properties can also be checked on the possible executions of a given pair of therapy and untimed CHA. Supervisory control for finite automata with CTL goals is known to be EXPTIME-complete, and controller synthesis algorithms exist [15].

The above representation of a cancer automaton is intuitive, but it does not include *timing*. It fails to model the fact that some transitions could be very short while others may take many years. In the next section we introduce timed CHAs, which are automata equipped with a set of real-valued variables, denoted as *clocks*, and constraints on the edges and states restrict the progression of the system. This model will be a special kind of hybrid automaton, justifying the word *hybrid* in ‘cancer hybrid automata’.

4 Timed CHAs

The framework we built so far is somewhat idealized in that transitions occur spontaneously and drugs can switch off transitions completely. More realistically, transitions would take certain durations of time, and drugs can slow down (or stop) the transition process. For example, in pancreatic cancer, it takes about a year for *K-ras* mutations in a cell to lead to neoplasms (so-called PanINs) [14]. To model durations, we will now add a notion of *time* to our CHA framework.

We start with the assumption that the acquisition of a hallmark requires a certain minimum amount of time. We do not specify exactly how that time is determined, but it could be the stopping time of a stochastic process such as randomizing over a set of driver mutations, or some value obtained from clinical data. Only after that time a given transition will be possible, and as mentioned, drugs can be used to prolong this time.

Further, we allow states to have *invariants*, specifying the maximum time that the system can remain in the respective state. For example, a tumor may only be able to remain in a state of unbounded growth without angiogenesis for a certain number of months.

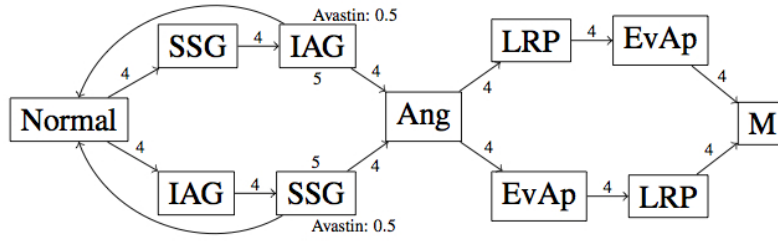


Figure 2: A simple timed CHA. The edges are labeled with the minimum times needed to make the respective transitions. In the two states that lead up to Angiogenesis, Avastin can be given to slow down the progress by half. Those states are labeled with invariants, and depending on the precise timing, these invariants can force the system back to Normal before the transition to Angiogenesis is possible.

Figure 2 shows the automaton from fig. 1 with timing information added, illustrating this intuition. We formalize the extension in the following.

We assume a finite set X of real-valued variables called *clocks*, over which the set of *constraints* $\mathcal{C}(X)$ is generated according to the grammar

$$\phi ::= x \geq k \mid \phi \wedge \phi ,$$

where $k \in \mathbb{N}$ and $x \in X$. A *valuation* of the variables in X is a mapping $\text{val} : X \rightarrow \mathbb{R}_{\geq 0}$. We denote the null valuation $x \mapsto 0$ by 0 . By $\text{val} \models \phi$ we denote that val satisfies ϕ .

Definition 1.1. A *timed CHA* is a tuple $H = (V, E, v_0, \ell, \rho)$ where

- V is a set of states,
- $E \subseteq V \times \mathcal{C}(X) \times V$ is a set of directed edges each labeled with a clock constraint,
- $v_0 \in V$ is the initial state,
- $\ell : V \times X \rightarrow \mathbb{N}$ is a partial function specifying the time limit (if any) for each clock that the system can remain in a given state (this is also called the invariant), and
- $\rho : V \times \mathcal{D} \times X \rightarrow \mathbb{R}_{\geq 0}$ yields a function specifying how a given drug influences the clocks at a given state.

Intuitively, at a given state v , the drug d modifies the clock rate, by slowing down or speeding up the clock x as specified by a multiplicative factor $\rho(v, d, x)$. When the factor is 1, the drug has no effect on that clock, and when it is 0, it effectively stops the clock from progressing. If several drugs have an effect on a clock, their factors are multiplied. We extend ρ to cocktails by setting $\rho(v, C, x) = \prod_{d \in C} \rho(v, d, x)$ for any cocktail $C \neq \emptyset$, and by convention, $\rho(v, \emptyset, x) = 1$.

A directed edge (v, ϕ, v') represents a transition from v to v' that can take place once the time constraint ϕ is satisfied.

We assume that for each state v that has a time limit for a clock x , there is an outgoing edge (v, ϕ, v') such that $\text{val} \models \phi$ for all val with $\text{val}(x) = \ell(v, x)$.³ This edge specifies the behavior of the system if the respective clock reaches its time limit.

³Note that this requires $\text{val} \models \phi$ even for valuations that exceed some other clock's invariant; however, this does not have an effect since we only allow \geq constraints on the edges.

The cost functions in the context of timed CHAs are the same as those for the untimed version, but with a timed interpretation: $c(v)$ is the cost of staying at state v per time unit (days/weeks/months/years), and $c(C)$ is the cost of administering a drug cocktail C per time unit.

We next see how to adapt the definitions related to runs of a CHA to the timed version, starting with the notion of a *timed state*.

Definition 1.2. A *timed state* of a timed CHA (V, E) is a tuple $(v, \text{val}) \in V \times \mathbb{R}^X$, where v is a state and val a clock valuation. There are two types of **transitions** between timed states:

1. Delay transitions, in symbols $(v, \text{val}) \xrightarrow{\delta, C} (v, \text{val}')$, where
 - $\delta \in \mathbb{R}_{>0}$ represents the (real) time delay,
 - C denotes the cocktail active during that time,
 - $\text{val}'(x) = \text{val}(x) + \delta \rho(v, C, x)$ for all x , and
 - $\text{val}'(x) \leq \ell(v, x)$ for all x with $\ell(v, x)$ defined.
2. State transitions, in symbols $(v, \text{val}) \rightarrow (v', 0)$, where
 - there is an edge $(v, \phi, v') \in E$ with $\text{val} \models \phi$.

Note that whenever a state transition takes place, the clocks are reset. This strategy simplifies our presentation and could be replaced by explicit clock resets as common in the literature.

This setup includes the special case where there is one clock unaffected by any drug, representing real time. Invariants over that clock can be used to specify, for example, the duration over which the tumor can remain in a certain state.

This timed setup can also emulate the concept of edges labeled with drugs that inhibit them. This model can be constructed as follows: Suppose we want to model an edge between two states v, v' that can be inhibited by a drug d . Then we can introduce a clock variable $x_{d,v'}$ with $\rho(v, d, x_{d,v'}) = 0$, and add a constraint $x_{d,v'} \geq z$ to the edge between v and v' , for some $z > 0$. As long as drug d is given before the constraint is satisfied, the transition will be inhibited. However, once the constraint is satisfied, the tumor has advanced too far and it is no longer possible to inhibit the transition.

A *run* in the case of a timed CHA H is a non-Zeno⁴ sequence of delay and state transitions. Similar as before, let $\text{Runs}((v, \text{val}), H)$ denote the set of runs that start in (v, val) . We write $\text{Runs}(H)$ for the set $\text{Runs}((v_0, 0), H)$, and $\text{Runs}_f((v, \text{val}), H)$ for the set of finite runs from $\text{Runs}((v, \text{val}), H)$.

Definition 1.3. A *therapy* is a function $\pi : \text{Runs}_f(H) \rightarrow 2^{\mathcal{D}}$. A *possible execution* of π in H is a run

$$S = (v_0, 0)(v_1, \text{val}_1)(v_2, \text{val}_2) \cdots$$

such that for all i with delay transitions $(v_i, \text{val}_i) \xrightarrow{\delta_i, C} (v_{i+1}, \text{val}_{i+1})$,⁵ for every $0 \leq \delta' < \delta_i$

$$\pi((v_0, 0) \dots (v_i, \text{val}_i)(v_i, \text{val}_i + \delta' \rho(v_i, C))) = C,$$

where $\rho(v_i, C)$ denotes the partial evaluation of ρ , i.e., the function $x \mapsto \rho(v_i, C, x)$.

This last condition ensures that the therapy does not change during a transition, or, put differently, that a change in therapy is always reflected by starting a new transition.

⁴That is, not containing an infinite chain of timed transitions with convergent total duration.

⁵Note that $v_i = v_{i+1}$.

For any finite run $r \in \text{Runs}_f(H)$, we denote its *duration* as

$$\tau(r) = \sum_{0 \leq j < \text{len}(r)} \begin{cases} \delta & \text{if } r_j \xrightarrow{\delta, C} r_{j+1} \text{ for some } \delta, C \\ 0 & \text{otherwise,} \end{cases}$$

where $\text{len}(r)$ denotes the length of the state sequence in r and r_i its initial segment of length i .

Definition 1.4. Given a CHA H and a possible execution S of a therapy π , the **cost** of S given π with discount factor $0 < d \leq 1$ is

$$c(S, \pi, H) = \sum_{i \geq 0} \frac{1}{d} \left(e^{-d\tau(S_i)} - e^{-d\tau(S_{i+1})} \right) (c(v_i) + c(\pi(S_i)))$$

(as before, by S_i we denote the initial segment of S up to step i). This simple discounting function does not necessarily capture a real patient's preferences, but any convergent function will work in its stead. We will consider more realistic functions in the future, which can potentially be designed on a case-by-case basis depending on the patient's valuation.

The set of possible costs of π in a timed CHA H is the set of costs of possible executions of π ,

$$c(\pi, H) = \{c(S, \pi, H) \mid S \text{ is possible execution of } \pi \text{ in } H\}.$$

The notions of Pareto dominance and universal therapies carry over from untimed CHAs.

4.1 Timed CTL

We can extend the CTL goals of the previous section to include time [2]. For example, the goal $\text{AG}_{\leq 20} \neg M$ says that metastasis is not reached within 20 time units (e.g., 20 years). This kind of goal represents the approach of turning cancer into a chronic disease, rather than trying to cure it completely. For example, the above formula may be appropriate for a patient of sixty years of age, who may then be able to get a less strenuous therapy, while for a younger patient the time requirements may be more extensive.

Out of all the therapies satisfying a CTL goal, the best ones may be chosen either by a separate cost optimization, or by incorporating cost requirements into the formulas using a weighted version of CTL [3].

5 Automatic therapy design for CHAs

Given the complexity of (timed) cancer progression and the influence of various drugs, the task of finding near-optimal therapy plans is (soon to be) beyond manual planning, and automated computational tools are very desirable.

The controller synthesis problem for different classes of automata have been studied in the literature, often restricted to achieving *safety* (avoiding a set of 'bad' states) and *reachability* (eventually reaching a 'goal' state) properties. Such properties form a sub-class of what can be expressed in richer temporal logics such as CTL. Safety properties are especially relevant for CHAs, because goals such as "metastasis will never be reached" can be expressed.

Untimed CHAs are a special kind of discrete automata for which efficient controller synthesis algorithms exist and can be applied to automatically design therapy-plans (see e.g. [18] for control using safety goals and [15] for an algorithm that uses CTL specifications).

Control of timed CHAs For timed CHAs, however, control is not as straightforward. CHAs are a special class of hybrid automata. Unfortunately, in hybrid systems, even simple verification and control problems like reachability and safety are undecidable [12]. However, several decidable subclasses of hybrid automata exist for which algorithms have been devised. One such subclass is that of *rectangular* hybrid automata. A rectangular automaton is an automaton in which the clock constraint on each edge is a rectangular region of continuous states. That is, it specifies for each clock a (possibly unbounded) interval that should contain its value. Also, the clock speed at each state is assumed to be bounded from below and above.

Rectangular automata form a most general class of hybrid automata for which even the reachability model checking problem is decidable [12, 10] and controller synthesis algorithms have been developed. For example, in [10] Henzinger et al. show that the control problem with LTL specifications is EXPTIME-complete in the size of the game, and 2EXPTIME-complete in the size of the formula.

These results rely on the requirement that the rectangular hybrid automata satisfies a property called *initialization* or *constant reset*. Initialization states that whenever the speed of a clock changes after a transition, the value of the variable is reinitialized to a fixed value (or a value in a fixed interval). This property cannot be relaxed without making the control problem undecidable [12]:

From timed CHAs to rectangular hybrid automata Timed CHAs bear a striking resemblance to rectangular hybrid automata, and it is thus worth exploring whether some of the controller synthesis results and algorithms can be applied to CHA models as well. Unfortunately, existing decidability results do not carry over directly because of some important differences between CHAs and (rectangular) hybrid automata.

First, in the hybrid automata literature, the rates of the clocks are generally assumed to be constant at any given state⁶ and what is controllable are (some of) the transitions between states. In the CHA framework, in contrast, the rates of the clocks is what can be affected by control actions (drugs), while the transitions (tumor progression) cannot be directly manipulated. However, this difference is mainly conceptual as a timed CHA can be translated to a hybrid automaton as follows:

Given a set of drugs \mathcal{D} and a CHA H with states V , we construct a hybrid automaton RH in the following way: For each state $v \in V$ and each cocktail $C \in 2^{\mathcal{D}}$, RH contains a state v_C with the same clock invariants as v . For any edge between two states $v, v' \in V$, RH contains an uncontrollable edge between v_C and v'_C , for each cocktail C , with the same clock constraints and resets as on the CHA edge. In addition to the uncontrollable edges, there are controllable directed edges from v_C to $v_{C'}$ for each v, C and C' . These edges represent changes of therapies, and have no clock constraints or resets. At a state v_C , the rate of each clock $x \in X$ is fixed, given by $\rho(v, C, x)$. This translation yields an automaton of size exponential in the number of drugs, but linear in the number of CHA states.

The result is a rectangular hybrid automaton. However, the translated CHA does not satisfy initialization, as the clock values (indicating progression time) are kept along controllable (change of cocktail) transitions while changing the rates of the clock. Thus, the negative results of Henzinger et al. [10] are no longer applicable.

Discretized control The simplest way around the undecidability of the control problem for rectangular hybrid automata that do not satisfy initialization is to allow for control moves (in our case, therapeutic interventions) only at discrete instants of time. Henzinger and Kopke [11] give an exponential-time algorithm for discrete-time safety control with CTL goals of rectangular hybrid automata with bounded

⁶One exception are so-called *differential games* [17], but their theory has not been well developed.

and non-decreasing variables. They also show the problem to be EXPTIME-hard and discrete-time verification of rectangular hybrid automata to be solvable in PSPACE.

Even though our definition of timed CHAs does not require clocks to be bounded, such a restriction would not impose a severe limitation. By bounding the clocks by some value that even the healthiest patient will never reach, we can thus aim for decidability without forfeiting any meaningful therapy. The algorithms from [11] do not directly apply to CHAs as their framework requires all discrete transitions to be controllable, whereas our cancer progression transitions are uncontrollable. However, they can be extended to include our framework via the following theorem, for which we only provide a sketch of the proof. The full proof can be found in the extended version of this paper.

Theorem 2 (Discrete control of bounded CHAs). *The controller synthesis problem of bounded discretized CHAs for CTL formulae can be solved in EXPTIME.*

Proof Sketch: First, we can translate the bounded CHA H into a rectangular hybrid automaton RH as described earlier. Then, the rectangular hybrid automaton RH can be described as a hybrid game⁷ HG by specifying that the controller is only allowed to make moves that include a change of therapy: from c to c' at state v by moving from (v, c) to (v, c') , and cancer is only allowed to pick an accessible new CHA state from the available $((v, c)(v', c))$ transitions.

Next we can extend the discretization method as given in [11] for rectangular automata to hybrid games. We can define a *sampling control game* DHG in which the players can only make one move every time unit, by adding a new variable x_{n+1} such that $\ell(v, x_{n+1}) = 1$ at each state; each clock constraint ϕ in the automaton becomes $\phi \wedge x_{n+1} \geq 1$ ($0 \leq x_{n+1} \leq 1$); the rate of x_{n+1} is 1 at all states; and x_{n+1} is reset to 0 after each discrete transition. This construction guarantees that moves by the cancer and therapist are always followed by a delay transition of duration 1⁸.

We can then define a bisimulation relation on the states of the discretized hybrid game DHG as in [11] as follows: We define an equivalence relation \approx_n on \mathbb{R}^n (the set of clock valuations) such that $\mathbf{y} \approx \mathbf{z}$ iff $\lfloor y_i \rfloor = \lfloor z_i \rfloor$ and $\lceil y_i \rceil = \lceil z_i \rceil$ for all $a \leq i \leq n$. Now, given two states (v, val) and (v', val') we define $(v, \text{val}) \cong_{DHG} (v', \text{val}')$ if $v = v'$ and $\text{val} \approx \text{val}'$. (we can also define $(v, \text{val}) \cong_{DHG}^m (v', \text{val}')$ for a bound m). We can show that this is indeed a bisimulation preserving CTL satisfaction, and since the result is a finite representation (exponential in size) of the original CHA, it follows that control of discretized bounded CHAs with CTL goals is solvable in EXPTIME. \square

6 Conclusions

This paper establishes a general formalism for describing cancer progression, without relying on any detailed mechanistic model of cancer pathways (which can be included independently as models of the discrete states). Our goal was to design a conceptually clear framework based on realistic biological foundations. As a case study, we have used this model to describe cancer hallmarks and their dynamics.

We discuss below how our framework can be used, as is, to model phenomena beyond what we discussed so far. Then, we point out the limitations of the current paper and give a list of topics that we plan to address in the near future.

⁷A game automaton is an automaton in which two players can make discrete moves. In our case not only the controller but also nature/the cancer can make discrete moves.

⁸Note, you have to assume that the automaton is big enough: in the original automaton it is not possible to make two moves in one time unit.

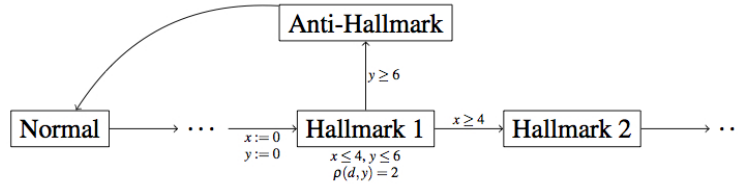


Figure 3: Illustrating how to model an anti-hallmark using two clocks x and y and a drug d that speeds up clock y at Hallmark 1 by a factor of 2.

6.1 Modeling growth, heterogeneity and anti-hallmarks

More general clocks: Thus far, we have referred to the clocks in CHAs as measuring *time*. However, they could be measuring different properties like *tumor size*, *motility* or *spatial properties*. For example, in the case of tumor size, the growth rate of the tumor may depend on the current discrete states of the progression and drugs can influence this rate. With this model we can reproduce the tumor growth dynamics as described in [22], by introducing two clocks: one measuring the number of stem cells and the other the number of differentiated cells. The various mutations can be modeled as transitions to a next state with different growth dynamics depending on the mutations already acquired.

Heterogeneity in tumors: So far we have modeled states of a CHA as representing the unique dominant phenotype of the tumor cell population. However, most forms of cancer are not likely to be monoclonal, i.e., consist of only one population in which the clonal expansions postulated by Hanahan and Weinberg take place, but rather involve several sub-populations of tumor cells [19], each with a distinct dominant phenotype [7, 13]. In order to model this heterogeneity, we can simply think of a CHA state as representing a *vector* of dominant phenotypes, one for each sub-population. One or several components of such a vector may differ from one state to the next, corresponding to a change of the dominant phenotype in the corresponding sub-population(s) during the respective transition; or the length of the vector may change, corresponding to new distinct sub-populations emerging or existing sub-populations dying out. This approach is, however, rather crude in modeling tumor heterogeneity, and does not straightforwardly accommodate, for example, information about tumor geometry or a model of the resulting spatial effects.

Anti-hallmarks Instead of trying to *slow down* cancer progression, there has recently been growing interest in approaches to *speed up* the process to a degree which will make the tumor nonviable and “push it over the edge” towards collapse. We refer to such nonviable states as *anti-hallmarks*. They can be modeled by putting constraints on the transitions leading to them that will never be satisfied, unless a drug is given which speeds up a certain clock. For example, consider the CHA in fig. 3. At Hallmark 1, without interference (both clocks increase with rate 1), the transition to Hallmark 2 will be taken after 4 time units. A drug that speeds up clock y by a factor of 2 will instead push the tumor to the Anti-Hallmark state, if given starting at most 1 time unit after entering Hallmark 1.

6.2 Extensions and Future Work

Partial observability and tests: The framework introduced in this paper assumes perfect information about the state of the system. In reality however, a clinician will only have partial observations of the

tumor's internal state. To reduce uncertainty about the current state of the cancer progression, *tests* can be performed. Our formal framework can be extended to include partial observability and tests, both for untimed and timed CHAs. Partial knowledge about the tumor's internal state can be modeled by introducing the notion of a belief set. Tests can be incorporated into the definition of a therapy as actions that reduce uncertainty about the current state. A therapy can then be described as a function from the set of belief-runs to cocktails or tests. The details appear in the full paper.

Compositional models: In a patient, cancer itself is not the only system of relevance. Other systems interact with the tumor's development, and especially during a therapeutic intervention, they need to be monitored. For example, the immune system and its role throughout carcinogenesis are receiving more and more attention [23], and the liver needs to be monitored to avoid damage due to excess toxicity. In principle, other subsystems of an organism could be modeled as hybrid automata in the same way as our CHA, which could then be composed to an overall model for which therapies with goals spanning all subsystems could be generated.

Building on our conceptual foundation, we plan to address several important issues next.

Algorithmic issues: In section 5, we have shown that the controller synthesis problem for timed CHAs is decidable if both the therapist and the cancer are only allowed to make moves at discrete moments in time. In the future, we plan to focus more on to the algorithmic side of verifying cancer hallmark automata, automatically generating therapies (including cost minimization), finding promising drug targets, etc.

Model extraction: Finally, we omitted a description of the methodologies needed for extracting cancer phenotypes and their temporal progression models from data or mechanistic pathway and population models. For example, there is currently no consensus that the cancer hallmarks described in the literature constitute a complete list, nor is there a clear understanding (either phenomenologically or mechanistically) of their precise discrete dynamics. We also believe that spatial structure (geometry, growth curve, spatial distribution of heterogeneity, etc.) as well as motility (self-seeding, circulating tumor cells) may hold additional and important clues that can be easily incorporated into our therapy design [6, 20]. Therefore, we plan to extract models from spatio-temporal data, for example, data obtained from detailed simulations, or gene expression and imaging data from patients or mouse models. We plan to use statistical inference algorithms for model extraction (such as GOALIE [21]) in order to reconstruct temporal (or spatio-temporal) phenomenological models of cancer-related processes from such data.

References

- [1] R.G. Abbott, S. Forrest & K.J. Pienta (2006): *Simulating the hallmarks of cancer*. *Artificial life* 12(4), pp. 617–634, doi:10.1162/artl.2006.12.4.617.
- [2] R. Alur, C. Courcoubetis & D. Dill (1993): *Model-Checking in Dense Real-Time*. *Information and Computation* 104(1), pp. 2 – 34, doi:10.1006/inco.1993.1024. Available at <http://www.sciencedirect.com/science/article/pii/S0890540183710242>.
- [3] Patricia Bouyer (2006): *Weighted Timed Automata: Model-Checking and Games*. *Electronic Notes in Theoretical Computer Science* 158(0), pp. 3–17, doi:10.1016/j.entcs.2006.04.002.

- [4] Edmund Clarke & E. Emerson (1982): *Design and synthesis of synchronization skeletons using branching time temporal logic*. In Dexter Kozen, editor: *Logics of Programs, Lecture Notes in Computer Science 131*, Springer Berlin / Heidelberg, pp. 52–71. Available at <http://dx.doi.org/10.1007/BFb0025774>.
- [5] Edmund M. Clarke, Orna Grumberg & Doron A. Peled (1999): *Model Checking*. MIT Press.
- [6] Elizabeth Comen, Larry Norton & Joan Massague (2011): *Clinical implications of cancer self-seeding*. *Nat Rev Clin Oncol* 8(6), pp. 369–377, doi:10.1038/nrclinonc.2011.64.
- [7] Isaiah J. Fidler (1978): *Tumor Heterogeneity and the Biology of Cancer Invasion and Metastasis*. *Cancer Research* 38(9), pp. 2651–2660. Available at <http://cancerres.aacrjournals.org/content/38/9/2651.abstract>.
- [8] Douglas Hanahan & Robert A. Weinberg (2000): *The Hallmarks of Cancer*. *Cell* 100(1), pp. 57–70, doi:10.1016/S0092-8674(00)81683-9.
- [9] Douglas Hanahan & Robert A. Weinberg (2011): *Hallmarks of Cancer: The Next Generation*. *Cell* 144(5), pp. 646–674, doi:10.1016/j.cell.2011.02.013. Available at <http://linkinghub.elsevier.com/retrieve/pii/S0092867411001279>.
- [10] Thomas A. Henzinger, Benjamin Horowitz & Rupak Majumdar (1999): *Rectangular Hybrid Games*. In: *In CONCUR 99, LNCS 1664*, Springer, pp. 320–335.
- [11] Thomas A. Henzinger & Peter W. Kopke (1999): *Discrete-time control for rectangular hybrid automata*. *Theoretical Computer Science* 221(1-2), pp. 369 – 392, doi:10.1016/S0304-3975(99)00038-9. Available at <http://www.sciencedirect.com/science/article/pii/S0304397599000389>.
- [12] Thomas A. Henzinger, Peter W. Kopke, Anuj Puri & Pravin Varaiya (1995): *What’s decidable about hybrid automata?* In: *Proceedings of the twenty-seventh annual ACM symposium on Theory of computing, STOC ’95*, ACM, New York, NY, USA, pp. 373–382.
- [13] Gloria H. Heppner (1984): *Tumor Heterogeneity*. *Cancer Research* 44(6), pp. 2259–2265. Available at <http://cancerres.aacrjournals.org/content/44/6/2259.short>.
- [14] Ralph H. Hruban, Michael Goggins, Jennifer Parsons & Scott E. Kern (2000): *Progression model for pancreatic cancer*. *Clinical cancer research* 6(8), p. 2969.
- [15] Shengbing Jiang & Ratnesh Kumar (2006): *Supervisory Control of Discrete Event Systems with CTL* Temporal Logic Specifications*. *SIAM Journal on Control and Optimization* 44(6), pp. 2079–2103, doi:10.1137/S0363012902409982. Available at <http://link.aip.org/link/?SJC/44/2079/1>.
- [16] Ji Luo, Nicole L. Solimini & Stephen J. Elledge (2009): *Principles of Cancer Therapy: Oncogene and Non-oncogene Addiction*. *Cell* 136(5), pp. 823–837, doi:10.1016/j.cell.2009.02.024. Available at <http://www.sciencedirect.com/science/article/B6WSN-4VS49KS-9/2/2dbe52f01a823f0c41169dbad5978c2f>.
- [17] Oded Maler (2002): *Control from computer science*. *Annual Reviews in Control* 26(2), p. 175–187.
- [18] Oded Maler, Amir Pnueli & Joseph Sifakis (1995): *On the synthesis of discrete controllers for timed systems*. In Ernst Mayr & Claude Puech, editors: *STACS 95, Lecture Notes in Computer Science 900*, Springer Berlin / Heidelberg, pp. 229–242.

- [19] Nicholas Navin, Jude Kendall, Jennifer Troge, Peter Andrews, Linda Rodgers, Jeanne McIndoo, Kerry Cook, Asya Stepansky, Dan Levy, Diane Esposito, Lakshmi Muthuswamy, Alex Krasnitz, W. Richard McCombie, James Hicks & Michael Wigler (2011): *Tumour evolution inferred by single-cell sequencing*. *Nature* 472(7341), pp. 90–94, doi:10.1038/nature09807.
- [20] Larry Norton (2008): *Cancer Stem Cells, Self-Seeding, and Decremental Exponential Growth: Theoretical and Clinical Implications*. *Breast Disease* 29(-1), pp. 27–36. Available at <http://iospress.metapress.com/content/T0R3351255U23705>.
- [21] Naren Ramakrishnan, Satish Tadepalli, Layne T. Watson, Richard F. Helm, Marco Antoniotti & Bud Mishra (2010): *Reverse engineering dynamic temporal models of biological processes and their relationships*. *Proceedings of the National Academy of Sciences*, doi:10.1073/pnas.1006283107. Available at <http://www.pnas.org/content/early/2010/06/16/1006283107.abstract>.
- [22] Ignacio Rodriguez-Brenes, Natalia Komarova & Dominik Wodarz (2011): *Evolutionary dynamics of feedback escape and the development of stem-cell-driven cancers*. *Proceedings of the National Academy of Sciences* 108(47).
- [23] Karin E. de Visser, Alexandra Eichten & Lisa M. Coussens (2006): *Paradoxical roles of the immune system during cancer development*. *Nature Reviews Cancer* 6(1), pp. 24–37, doi:10.1038/nrc1782.